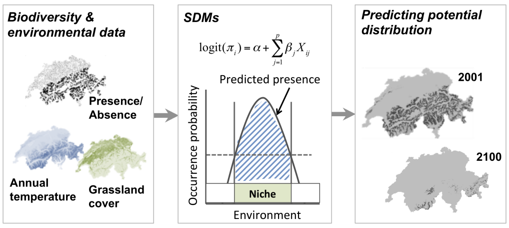
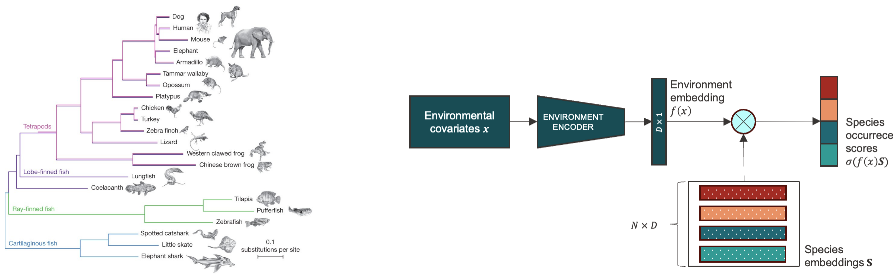
Master Thesis Project:
Integrating phylogeny in to species distribution models

Species distribution models (SDMs) relate species occurrence data with environmental variables, such as climate, soil, topography, satellite images, etc. They are used to understand and predict the distributions of species across landscapes and are of particular interest for ecosystem conservation. Statistical models and machine learning models have become standard for SDMs and several recent studies have shown promising results in using deep learning. However, deep learning may have untapped potential to address the limitations of current methods. In particular, neural networks present a flexible framework allowing the integration of domain knowledge that may be valuable in constructing more ecologically realistic SDMs.



SDM illustration from `https://damariszurell.github.io/EEC-MGC/b3_SDM_intro.html`

This project will focus on integrating phylogenetic information into SDMs. Phylogenies, or phylogenetic trees, depict the evolutionary relationships between biological species, based on genetic similarities. When considering species' distributions through an evolutionary perspective, the idea is that closely related species are ecologically similar and will respond similarly to environmental variables. Here, the aim is to constrain SDM predictions by learning species embeddings that are more similar for species that are phylogenetically closer and vice-versa. A neural network (MLP or CNN) will be used as multi-label classifier to model multiple species simultaneously. The vectors of the weight matrix of the final fully connected layer, which performs the classification, can be interpreted as species embeddings (see illustration below). For instance, constrastive learning may be used to push species embeddings of similar species closer in the feature space, and those of dissimilar species further apart. Other methods may also be explored to constrain these species embeddings according to phylogenetic similarity.



Example of a phylogenetic tree (left), SDM framework with species embeddings (right)

ECEO

Nina van Tiel
EPFL Valais Wallis
ENAC-IIE-ECEO
Rue des Ronquos 86
1951 Sion
nina.vantiel@epfl.ch
https://people.epfl.ch/nina.vantiel

This project will use occurrence data of species across Switzerland, with most likely a focus on plant species.

Requirements

- Proficiency in python
- Experience in deep learning
- Interest in biology and/or ecology
- Strong willingness to learn and ability to work independently

Contact
Nina van Tiel (nina.vantiel@epfl.ch)

References

- S. Beery, E. Cole, J. Parker, P. Perona, and K. Winner. Species distribution modeling for machine learning practitioners: A review. In ACM SIGCAS conference on computing and sustainable societies, pages 329–348, 2021.
- J. B. Losos. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecology letters, 11(10):995–1003, 2008.
- D. Chen, Y. Xue, S. Chen, D. Fink, and C. Gomes. Deep multi-species embedding. arXiv preprint arXiv:1609.09353, 2016.
- O. Mac Aodha, E. Cole, and P. Perona. Presence-only geographical priors for fine-grained image classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9596–9606, 2019.
- P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. Supervised contrastive learning. Advances in neural information processing systems, 33: 18661–18673, 2020.

ECEO

Nina van Tiel
EPFL Valais Wallis
ENAC-IIE-ECEO
Rue des Ronquos 86
1951 Sion

nina.vantiel@epfl.ch
https://people.epfl.ch/nina.vantiel